INTRODUCTION TO

CAUSAL MODELLING AND REASONING

Martina Cinquini & Isacco Beretta
CORRELATION DOES NOT IMPLY CAUSATION

Number of people who drowned by falling into a pool correlates with Films Nicolas Cage appeared in

Correlation: 66.6% (r=0.666004)
CORRELATION DOES NOT IMPLY CAUSATION

Sleeping with shoes on is strongly correlated with waking up with a headache
CORRELATION DOES NOT IMPLY CAUSATION

Sleeping with shoes on is strongly correlated with waking up with a headache.

Common cause: drinking the night before.

1. Shoe-sleepers differ from non-shoe-sleepers in a key way.
CORRELATION DOES NOT IMPLY CAUSATION

Sleeping with shoes on is strongly correlated with waking up with a headache

Common cause: drinking the night before

1. Shoe-sleepers differ from non-shoe-sleepers in a key way
2. Confounding
CORRELATION DOES NOT IMPLY CAUSATION

Sleeping with shoes on is strongly correlated with waking up with a headache

Common cause: drinking the night before

1. Shoe-sleepers differ from non-shoe-sleepers in a key way
2. Confounding

Total association (e.g., correlation):
Mixture of causal and confounding association
INGREDIENTS OF A STATISTICAL THEORY OF CAUSALITY

- Working definition of causation
- Method for creating causal models
- Method for linking causal models with features of data
- Method for reasoning over model and data
THE LATTER OF CAUSALITY

"Actual" Causality

“Causality-in-mean”

Statistics

3. COUNTERFACTUALS
ACTIVITY: Imagining, Retrospection, Understanding
QUESTIONS: What if I had done...? Why?
(Was it X that caused Y? What if X had not
occurred? What if I had acted differently?)
EXAMPLES: Was it the aspirin that stopped my headache?
Would Kennedy be alive if Oswald had not
killed him? What if I had not smoked for the
last 2 years?

2. INTERVENTION
ACTIVITY: Doing, Intervening
QUESTIONS: What if I do...? How?
(What would Y be if I do X?
How can I make Y happen?)
EXAMPLES: If I take aspirin, will my headache be cured?
What if we ban cigarettes?

1. ASSOCIATION
ACTIVITY: Seeing, Observing
QUESTIONS: What if I see...?
(How are the variables related?
How would seeing X change my belief in Y?)
EXAMPLES: What does a symptom tell me about a disease?
What does a survey tell us about the
election results?
RANDOMIZED EXPERIMENTS

Which kind of post works better?

Interventional data

Post 1

Post 2

Post ...

Post n

Limitations

- Can not use historical data
- It cannot be applied to certain situations (e.g., long-term effect, selected demographics, content virality)
BEYOND RANDOMIZED EXPERIMENTS

Associational data

Individual who decides where the post are sent

Historical Data

Post 1

Post 2

Post ...

Post n

New Data
CAUSAL MODEL FRAMEWORKS

Potential Outcomes (PO)  Structural Causal Model (SCM)

Antecedents in the earlier econometric literature

Demand and Supply Models (Haavelmo, 1944)

Path analysis (Wright, 1934)

These frameworks are complementary, with different strengths that make them appropriate to address different problems in specific situations.
CAUSAL MODEL FRAMEWORKS

Potential Outcomes (PO)

Antecedents in the earlier econometric literature

Demand and Supply Models (Haavelmo, 1944)

Specifically, to deal with:

Estimating individual-level causal effects

Structural Causal Model (SCM)

Path analysis (Wright, 1934)

Complex models with a large number of variables
POTENTIAL OUTCOME: INTUITION

Inferring the effect of treatment on some outcome
Inferring the effect of treatment on some outcome

Causal Effect?

POTENTIAL OUTCOME: INTUITION

Take a pill

Don’t take a pill
POTENTIAL OUTCOME: INTUITION

Inferring the effect of treatment on some outcome

No Causal Effect

Take a pill

Don’t take a pill
POTENTIAL OUTCOME: NOTATION

\[
do (T = 1) \quad Y_i|\text{do}(T = 1) = Y_i(1)
\]

\[
do (T = 0) \quad Y_i|\text{do}(T = 0) = Y_i(0)
\]

- **T**: Observed Treatment
- **Y**: Observed Outcome
- **i**: used in subscript to denote a specific individual
- **Y_i(1)**: PO under treatment
- **Y_i(0)**: PO under no treatment
OTHER DEFINITIONS

\[ Y = \text{unit (individual)} \]

\[ \text{population} \]

\[ \text{Age, Gender, Weight} = \text{covariates of the individual (Z)} \]

**INDIVIDUAL TREATMENT EFFECT (ITE)**

The ITE for the \( i \)th unit is defined as follows:

\[ Y_i(1) - Y_i(0) \]
POTENTIAL OUTCOME: NOTATION

\[
\begin{align*}
\text{do (T = 1)} & \quad Y_i(1) = 1 \\
\text{do (T = 0)} & \quad Y_i(0) = 0
\end{align*}
\]

- \(T\): Observed Treatment
- \(Y\): Observed Outcome
- \(i\): used in subscript to denote a specific individual
- \(Y_i(1)\): PO under treatment
- \(Y_i(0)\): PO under no treatment

\textbf{Causal Effect}: \(Y_i(1) - Y_i(0) = 1\)
The fundamental problem of causal inference

We cannot observe both $Y_i(1)$ and $Y_i(0)$, therefore we cannot observe the Causal Effect: $Y_i(1) - Y_i(0)$.

The PO that you do not (and cannot) observe are known as COUNTERFACTUALS because they are counter to fact (reality).

Due to the fundamental problem, we know that we can’t access to ITE.
The ATE is obtained by taking an average over the ITEs:

$$E[Y_i(1) - Y_i(0)] = E[Y(1) - Y(0)]$$

where we recall that the average is over the individuals $i$ if $Y_i(x)$ is deterministic.

How would we actually compute the ATE?
The fundamental problem of CI can be seen as a **MISSING DATA PROBLEM**.

The question mark means that we do not observe the value.
The fundamental problem of CI can be seen as a **MISSING DATA PROBLEM**

\[ \text{E}[Y_i(1) - Y_i(0)] = ? \]

<table>
<thead>
<tr>
<th>(i)</th>
<th>(T)</th>
<th>(Y)</th>
<th>(Y(1))</th>
<th>(Y(0))</th>
<th>(Y(1) - Y(0))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td>0</td>
<td>?</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td>0</td>
<td>?</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>?</td>
<td>1</td>
<td>?</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>
THE FUNDAMENTAL PROBLEM OF CAUSAL INERENCE

\[ E[Y_i(1) - Y_i(0)] = E[Y(1)] - E[Y(0)] \]

<table>
<thead>
<tr>
<th>( i )</th>
<th>( T )</th>
<th>( Y )</th>
<th>( Y(1) )</th>
<th>( Y(0) )</th>
<th>( Y(1) - Y(0) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td>0</td>
<td>?</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td>0</td>
<td>?</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>?</td>
<td>1</td>
<td>?</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

The fundamental problem of CI can be seen as a **MISSING DATA PROBLEM**
THE FUNDAMENTAL PROBLEM OF CAUSAL INFERENCE

\[ E[Y_i(1) - Y_i(0)] = E[Y(1)] - E[Y(0)] = E[Y \mid T = 1] - E[Y \mid T = 0] \]

<table>
<thead>
<tr>
<th>(i)</th>
<th>(T)</th>
<th>(Y)</th>
<th>(Y(1))</th>
<th>(Y(0))</th>
<th>(Y(1) - Y(0))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td>0</td>
<td>?</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td>0</td>
<td>?</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>?</td>
<td>1</td>
<td>?</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

The fundamental problem of CI can be seen as a **MISSING DATA PROBLEM**
THE FUNDAMENTAL PROBLEM OF CAUSAL INFERENCE

The fundamental problem of CI can be seen as a **MISSING DATA PROBLEM**

\[
E[Y_i(1) - Y_i(0)] = E[Y(1)] - E[Y(0)] = E[Y \mid T = 1] - E[Y \mid T = 0]
\]

<table>
<thead>
<tr>
<th>i</th>
<th>T</th>
<th>Y</th>
<th>Y(1)</th>
<th>Y(0)</th>
<th>Y(1)−Y(0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>?</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>?</td>
<td></td>
</tr>
</tbody>
</table>
## THE FUNDAMENTAL PROBLEM OF CAUSAL INFERENCE

The fundamental problem of CI can be seen as a **MISSING DATA PROBLEM**

\[
E[Y_i(1) - Y_i(0)] = E[Y(1)] - E[Y(0)] = E[Y | T = 1] - E[Y | T = 0]
\]

<table>
<thead>
<tr>
<th>(i)</th>
<th>(T)</th>
<th>(Y)</th>
<th>(Y(1))</th>
<th>(Y(0))</th>
<th>(Y(1) - Y(0))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>?</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>?</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>?</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>?</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>?</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>?</td>
</tr>
</tbody>
</table>

\[
\frac{2}{3} \quad \frac{1}{3}
\]

The fundamental problem of CI can be seen as a **MISSING DATA PROBLEM**.
The fundamental problem of CI can be seen as a **missing data problem**.

The table shows the values of $Y_i(1)$, $Y_i(0)$, and the difference $Y_i(1) - Y_i(0)$ for different values of $i$.

The fundamental problem of CI can be seen as a **missing data problem**.
## THE FUNDAMENTAL PROBLEM OF CAUSAL INFERENCE

\[ E[Y_i(1) - Y_i(0)] = E[Y(1)] - E[Y(0)] \neq E[Y | T = 1] - E[Y | T = 0] \]

<table>
<thead>
<tr>
<th>( i )</th>
<th>( T )</th>
<th>( Y )</th>
<th>( Y(1) )</th>
<th>( Y(0) )</th>
<th>( Y(1) - Y(0) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>?</td>
<td></td>
</tr>
</tbody>
</table>

What does it mean?  
**causation is simply association**

In general, they are not equal due to **CONFOUNDING**

What **ASSUMPTIONS** would make the ATE equal to the associational difference?
**IGNORABILITY - \((Y(1), Y(0)) \perp T\)**

\[
E[Y_i(1)] - E[Y_i(0)] = E[Y(1) \mid T = 1] - E[Y(0) \mid T = 0]
\]

\[
= E[Y \mid T = 1] - E[Y \mid T = 0]
\]

- We can ignore how individual ended up in the treatment/control group, and treat their PO as **exchangeable**. However, it is **unrealistic** in observational data.

- **Unconfoundeness**

\[
(Y(1), Y(0)) \perp T \mid X
\]
IGNORABILITY - \((Y(1), Y(0)) \perp T\)

\[
E[Y_i(1)] - E[Y_i(0)] = E[Y|T = 1] - E[Y|T = 0]
\]

\[
= E[Y|T = 1] - E[Y|T = 0]
\]

- We can ignore how individual ended up in the treatment/control group, and treat their PO as exchangeable. However, it is **unrealistic** in observational data.

- **Unconfoundeness**

  \((Y(1), Y(0)) \perp T | X\)

  When conditioning on \(X\), **non-causal** association between \(T\) and \(Y\) no longer exists.
While is not a problem in randomized experiments, it is an **untestable assumption** in observational data.

There may be some **unobserved confounders** that are not part of $X = \{M\}$, meaning unconfoundedness is **violated**.

**Ignorability**

$$(Y(1), Y(0)) \perp T \mid X$$
While is not a problem in randomized experiments, it is an **untestable assumption** in observational data.

There may be some **unobserved confounders** that are not part of $X = \{M\}$, meaning unconfoundedness is **violated**.

Ignorability 

$$(Y(1), Y(0)) \perp T | X$$
ASIDE: IDENTIFIABILITY

\[ E[Y_i(1)] - E[Y_i(0)] = E[Y(1) \mid T = 1] - E[Y(0) \mid T = 0] \]

\[ = E[Y \mid T = 1] - E[Y \mid T = 0] \]

Causal quantities

Statistical quantities

A causal quantity (e.g. \( E[Y(t)] \)) is **identifiable** if we can compute it from a purely statistical quantity (e.g. \( E[Y \mid t] \))
For all values $x$ of covariates $x$ present in the population of interest (i.e., $z$ such that $P(X = x > 0)$)

$$0 < P(T = 1 | X = x) < 1$$

Positivity is the condition that **all subgroups of the data** with different value $x$ for covariates $X$ have some probability of receiving any value of treatment $T$
POSITIVITY: INTUITION

Total Population

X = x

No one treated

T = 0
T = 0  T = 0
T = 0

Everyone treated

T = 1
T = 1  T = 1
T = 1
POSITIVITY: OVERLAP

No overlap means severe positivity violation.

Complete overlap means no positivity violation.
adjusting (conditioning) on more covariates $Z$

could lead to

higher chance of satisfying unconfoundedness

could lead to

higher chance of violating positivity

demanding too much from models and getting very bad behavior in return

fit a model to $\mathbb{E}[Y|X,Z]$ using the available data $(x,y,z)$

increase the "dimension" of the covariates $Z$

makes the subgroups for any level $z$ of the covariates $Z$ smaller

CURSE OF DIMENSIONALITY
The outcome $Y_i$ of each unit $i$ is unaffected by anyone else's treatment $T_j$ $j \neq i$

$$Y_i(t_1, t_2, ..., t_{i-1}, t_{i+1}, ..., t_{n-1}, t_n) = Y_i(t_i)$$
My happiness
NO INTERFERENCE

Whether friends get dogs

$T_1 \quad \ldots \quad T_{i-1} \quad T_i \quad T_{i+1} \quad \ldots \quad T_n$

My happiness

Whether friends get dogs
If the treatment is $T$, then the observed outcome $Y$ is the potential outcome under treatment $X$.

Formally, $T = t \iff Y = Y(t)$
SUTVA

A combination of consistency and no interference. Specifically, the PO of a unit do not depend on the treatments assigned to others.

But in real world ...

\[(T = 1) \implies Y = 1 \text{ (I’m happy)}\]
It match T=0 and T=1 observations on the estimated probability of being treated.
HOW TO USE THE PO: AN EXAMPLE

PROPENSITY SCORE MATCHING (PSM)

It match $T=0$ and $T=1$ observations on the estimated probability of being treated.

![Diagram of propensity score matching](image)
PO RECAP

- Mainly used for estimating average effects of binary treatments
- Convincing empirical applications

LIMITATIONS:

- An expert of the field should verify whether all the previous assumptions are valid.
  It is challenging and you need some people working on it.
- No use of causal diagrams
CAUSAL MODEL FRAMEWORKS

Potential Outcomes (PO)

Demand and Supply Models
(Haavelmo, 1944)

Structural Causal Model (SCM)

Path analysis
(Wright, 1934)

Antecedents in the earlier econometric literature

Specifically, to deal with:

Estimating individual-level causal effects

Complex models with a large number of variables
Mathematically, a Structural Causal Model (SCM) consists of a set of Endogenous (V) and a set of Exogenous (U) variables connected by a set of functions (F) that determine the values of the variables in V based on the values of the variables in U.

Each SCM is associated with a graphical model where each node is a variable in V and each edge is a function $f$. 
GRAPH TERMINOLOGY

Directed Graph

Undirected Graph
This graph contains a cycle
GRAPH TERMINOLOGY

- Directed Acyclic Graph
- Undirected Graph
Descendant is a **broader** term than child because it includes **not only the immediate children** but also their children and so forth.
Adjacent is a node that is **directly connected** to another node within a graph.
A **path** is a sequence of nodes where each node is connected to the next node by an edge.
**STRUCTURAL CAUSAL MODEL: EXAMPLE**

\[
X = \{X_1, X_2, X_3\}
\]

\[
X_1 := \text{Uniform}(0, 1)
\]

\[
X_2 := \sin(X_1) + \text{Normal}(0, 1)
\]

\[
X_3 := 2 \times X_1 + \text{Normal}(0, 1)
\]

**Structural Equation**

(SE)

**Directed Acyclic Graph**

(DAG)
CAUSAL STRUCTURES

Chain

Collider

Confounder
CAUSAL STRUCTURES: EXAMPLE

Sprinkler $X_3$ affects Season $X_1$ and Wet $X_4$. Rain $X_2$ affects Wet $X_4$. Wet $X_4$ affects Slippery $X_5$.
CAUSAL STRUCTURES: EXAMPLE

Confounder
CAUSAL STRUCTURES: EXAMPLE

Collider
LEVELS OF INVESTIGATION

Causal Discovery (CD)
Given a set of variables, is it possible to **determine the causal relationship** between them?

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2</td>
<td>2.2</td>
<td>1.6</td>
<td>7.5</td>
<td>2.4</td>
</tr>
<tr>
<td>2.9</td>
<td>3.1</td>
<td>1.3</td>
<td>8.2</td>
<td>5.1</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Causal Inference (CI)
If we manipulate the value of one variable, how much would the others change?

?
CAUSAL PIPELINE

Dataset

Causal Discovery

What are the consequences of turning on the sprinkler?
(The floor gets wet)

Causal Inference
CAUSAL DISCOVERY: METHODS

Constraint-based

Score-based

Markov Equivalence Class

\[ X_1 \perp \! \! \perp X_3 \mid X_2 \text{ and } X_1 \not\perp \! \! \perp X_3 \]
CAUSAL DISCOVERY: METHODS

Constraint-based

Score-based

Markov Equivalence Class

\( X_1 \independent X_3 \mid X_2 \) and \( X_1 \not\independent X_3 \)

V-structure

\( X_1 \not\independent X_3 \mid X_2, \quad X_1 \independent X_3 \)
CAUSAL DISCOVERY: METHODS

- **Constraint-based**
- **Score-based**

**Functional Causal Models**

Markov Equivalence Class

\[ X_1 \perp X_3 \mid X_2 \quad \text{and} \quad X_1 \not\perp X_3 \]

- **Strong** assumptions but they **can uniquely identify the true DAG**
- Linear and non-Gaussian, Additive noise, Post-nonlinear
**INTERVENTION**

Interpreting edges as cause-effect relationships enable reasoning about the outcome of interventions using the **do-operator**.
The notation $\text{do}(\text{Sprinkler} := \text{ON})$ denotes an intervention by which variable Sprinkler is set to value ON.

Externally forcing the variable to assume a particular value makes it independent of its causes and breaks their causal influence on it.
Graphically, the effect of an intervention can be captured by removing all incoming edges to the intervened variable.
The best-known technique to find causal estimands given a graph.

A set of variables $Z$ satisfies the **back-door criterion** relative to an ordered pair of variables $(X_i, X_j)$ in a DAG $G$ if:

- no node in $Z$ is a descendant of $X_i$
- $Z$ blocks every path between $X_i$ and $X_j$ that contains an arrow into $X_i$. 

**BACK-DOOR CRITERION**
This path is not causal. It is a process that creates spurious correlations between $X_1$ and $X_3$ that are driven solely by fluctuations in the $X_2$ random variable.

If we can close all of the open backdoor paths, then we can isolate the causal effect of $X_1$ and $X_3$ using an identification strategy.

$$P(X_3 \mid do(X_1)) = \sum X_2 P(X_3 \mid X_1, X_2) P(X_2)$$
EXERCISE

Find the discovered graph


https://www.bradyneal.com/causal-inference-course
THANK FOR YOUR ATTENTION